

ABSTRACT OF THE DISCLOSURE

a¹

A technique for generating cross-references among categories in a knowledge base extracts a plurality of themes from a corpus of documents. A theme identifies subject matter contained in a corresponding document. A plurality of scores are generated such that each score identifies a relative theme strength among theme pairs of the themes extracted from the documents. In general, a theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in the document. Thereafter, the most related theme pairs are selected as indicated by the scores. Category pairs of the knowledge base are then selected by mapping the themes of the selected theme pairs to corresponding categories of the knowledge base. A cross-reference between categories of the category pairs in the knowledge base is generated so as to identify an association between the category pairs.

IN THE CLAIMS:

Please amend the claims 1-3, 6-8, and 11-13 as follows:

1. (Once Amended) A method for generating cross-references among categories in a knowledge base, said method comprising the steps of:

a²

extracting, from a plurality of documents, a plurality of themes, wherein a theme identifies subject matter contained in a corresponding document;

generating a theme strength for said themes, said theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in said document;

generating a plurality of scores, from said theme strengths, to identify a relative theme pair strength for at least one pair of said themes extracted from said documents;

selecting theme pairs based on said scores;

selecting category pairs in said knowledge base by mapping said themes of said theme pairs selected to corresponding categories of said knowledge base; and

Q² generating a cross reference in said knowledge base between categories of said category pairs, wherein said cross reference identifies an association between said category pairs.

2. (Once Amended) The method as set forth in claim 1, wherein the step of generating a plurality of scores comprises the steps of:

generating a matrix comprising a plurality of columns and rows to form a plurality of entries, wherein each column represents one of said themes and each row represents one of said themes; and

generating a score for at least a subset of said entries of said matrix, such that a score reflects a relative theme pair strength between two themes represented by said entry for said documents.

3. (Once Amended) The method as set forth in claim 2, wherein the step of generating a score for at least a subset of said entries of said matrix comprises the steps of:

a² calculating a plurality of products for an entry by multiplying theme strengths corresponding to two themes represented by said entry for each document that includes said two themes represented by said entry; and

summing said products for an entry to generate said score.

6. (Once Amended) A system comprising:

search and retrieval module for receiving a user query and for generating a query response including query feedback;

a knowledge base, coupled to said search and retrieval module, for storing relationships among terminology for use as query feedback;

a³ a knowledge base processing system, coupled to said knowledge base for processing a plurality of documents and automatically extending said relationships among said terminology in said knowledge base, said knowledge base processing system for extracting, from said documents, a plurality of themes, wherein a theme identifies subject matter contained in a corresponding document, for generating a theme strength for said themes, said theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in said document, for generating a plurality of scores, from said theme strengths, to identify a relative theme pair strength for at least one pair of said themes extracted from said documents, for selecting theme pairs

based on said scores, for selecting category pairs in said knowledge base by mapping said themes of said theme pairs selected to corresponding categories of said knowledge base, and for generating a cross reference in said knowledge base between categories of said category pairs, wherein said cross reference identifies an association between said category pairs.

a³

7. (Once Amended) The system as set forth in claim 6, wherein the knowledge base processing system further for generating a matrix comprising a plurality of columns and rows to form a plurality of entries, wherein each column represents one of said themes and each row represents one of said themes and for generating a score for at least a subset of said entries of said matrix, such that a score reflects a relative theme pair strength between two themes represented by said entry for said documents.

8. (Once Amended) The system as set forth in claim 7, wherein the knowledge base processing system further for calculating a plurality of products for an entry by multiplying theme strengths corresponding to two themes represented by said entry for each document that includes said two themes represented by said entry, and for summing said products for an entry to generate said score.

a⁴

11. (Once Amended) A computer readable medium comprising a plurality of instructions, which when executed, causes the computer to perform the steps of:

extracting, from a plurality of documents, a plurality of themes, wherein a theme identifies subject matter contained in a corresponding document;

generating a theme strength for said themes, said theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in said document;

generating a plurality of scores, from said theme strengths, to identify a relative theme pair strength for at least one pair of said themes extracted from said documents;

selecting theme pairs based on said scores;

a⁴ selecting category pairs in said knowledge base by mapping said themes of said theme pairs selected to corresponding categories of said knowledge base; and

generating a cross reference in said knowledge base between categories of said category pairs, wherein said cross reference identifies an association between said category pairs.

12. (Once Amended) The computer readable medium as set forth in claim 11, wherein the step of generating a plurality of scores comprises the steps of:

generating a matrix comprising a plurality of columns and rows to form a plurality of entries, wherein each column represents one of said themes and each row represents one of said themes; and